Performance of U.S. 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context

First Look at PISA 2012

NCES 2014-024

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Performance of U.S. 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context

First Look at PISA 2012

DECEMBER 2013

Dana Kelly
National Center for Education Statistics

Holly Xie
Project Officer
National Center for Education Statistics

Christine Winquist Nord
Frank Jenkins
Jessica Ying Chan
David Kastberg
Westat

NCES 2014-024
U.S. DEPARTMENT OF EDUCATION

NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Jack Buckley
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

NCES, IES, U.S. Department of Education
1990 K Street, NW
Washington, DC 20006-5651

December 2013

The NCES Home Page address is http://nces.ed.gov.
The NCES Publications and Products address is http://nces.ed.gov/pubsearch.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-10-C-0047 with Westat. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Kelly, D., Xie, H., Nord, C.W., Jenkins, F., Chan, J.Y., and Kastberg, D. (2013). *Performance of U.S. 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context: First Look at PISA 2012* (NCES 2014-024). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

Content Contact
Holly Xie
(202) 502-7314
holly.xie@ed.gov

Acknowledgments

The authors wish to thank the students, teachers, and school officials who participated in the Program for International Student Assessment (PISA) 2012. Without their assistance and cooperation, this study would not be possible. The authors also wish to thank all those who contributed to the PISA design, implementation, and data collection as well as the writing, production, and review of this report.

This page intentionally left blank

Contents

List of Tables

List of Figures

List of Exhibits

Introduction

What Is PISA?

The Program for International Student Assessment (PISA) is a system of international assessments that allows countries to compare outcomes of learning as students near the end of compulsory schooling. PISA core assessments measure the performance of 15-year-old students in mathematics, science, and reading literacy every 3 years. Coordinated by the Organization for Economic Cooperation and Development (OECD), PISA was first implemented in 2000 in 32 countries. It has since grown to 65 education systems in 2012.

What PISA Measures

PISA's goal is to assess students' preparation for the challenges of life as young adults. PISA assesses the application of knowledge in mathematics, science, and reading literacy to problems within a real-life context (OECD 1999). PISA does not focus explicitly on curricular outcomes and uses the term "literacy" in each subject area to indicate its broad focus on the application of knowledge and skills. For example, when assessing mathematics, PISA examines how well 15-year-old students can understand, use, and reflect on mathematics for a variety of real-life problems and settings that they may not encounter in the classroom. Scores on the PISA scales represent skill levels along a continuum of literacy skills.

Each PISA data collection cycle assesses one of the three core subject areas in depth (considered the major subject area), although all three core subjects are assessed in each cycle (the other two subjects are considered minor subject areas for that assessment year). Assessing all three subjects every 3 years allows countries to have a consistent source of achievement data in each of the three subjects, while rotating one area as the primary focus over the years. Mathematics was the major subject area in 2012, as it was in 2003, since each subject is a major subject area once every three cycles. In 2012, mathematics, science, and reading literacy were assessed primarily through a paper-and-pencil assessment, and problem solving was administered via a computer-based assessment. In addition to these core assessments, education systems could participate in optional paper-based financial literacy and computer-based mathematics and reading assessments. The United States participated in these optional assessments. Visit www.nces.ed.gov/surveys/pisa for more information on the PISA assessments, including information on how the assessments were designed and examples of PISA questions.

Mathematics Literacy

In PISA 2012, the major subject was mathematics literacy, defined as:

> *An individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged, and reflective citizens* (OECD 2013, p. 25).

More specifically, the PISA mathematics assessment looks at four mathematical content categories and three mathematical process categories:[1]

Mathematical content categories (OECD 2013, pp. 33–35):

- **Change and relationship:** Can students model change and relationships with the appropriate functions and equations?
- **Space and shape:** Can students understand perspective, create and read maps, and manipulate 3D objects?
- **Quantity:** Are 15-year-olds able to comprehend multiple representations of numbers, engage in mental calculation, employ estimation, and assess the reasonableness of results?
- **Uncertainty and data:** Can students use probability and statistics and other techniques of data representation and description to mathematically describe, model, and interpret uncertainty?

Mathematical process categories (OECD 2013, pp. 28–30):

- **Formulate:** Can 15-year-olds recognize and identify opportunities to use mathematics and then provide mathematical structure to a problem presented in some contextualized form in order to formulate situations mathematically?
- **Employ:** Are students able to employ mathematical concepts, facts, procedures, and reasoning to solve mathematically formulated problems and obtain mathematical conclusions?
- **Interpret:** Can students interpret, apply, and evaluate mathematical outcomes in order to determine whether results are reasonable and make sense in the context of the problem?

The PISA mathematics framework was updated for the 2012 assessment. The revised framework is intended to clarify the mathematics relevant to 15-year-old students, while ensuring that the items developed remain set in meaningful and authentic contexts, and defines the mathematical processes in which students engage as they solve problems. These processes, described above, are being used for the first time in 2012 as a primary reporting dimension. Although the framework has been updated, it is still possible to measure trends in mathematics literacy over time, as the underlying construct is intact.

Mathematics literacy is reported both in terms of proficiency levels and scale scores (reported on a scale of 0–1,000). Exhibit 1 (see following page) describes the six mathematics literacy proficiency levels and their respective cut scores.

[1] Mathematics literacy subscale results can be found online at http://nces.ed.gov/surveys/pisa/pisa2012/.

Exhibit 1. Description of PISA proficiency levels on mathematics literacy scale: 2012

Proficiency level and lower cut score	Task descriptions
Level 6 669	At level 6, students can conceptualize, generalize, and utilize information based on their investigations and modeling of complex problem situations, and can use their knowledge in relatively non-standard contexts. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understanding, along with a mastery of symbolic and formal mathematical operations and relationships, to develop new approaches and strategies for attacking novel situations. Students at this level can reflect on their actions, and can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments and the appropriateness of these to the original situations.
Level 5 607	At level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem-solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterizations, and insight pertaining to these situations. They begin to reflect on their work and can formulate and communicate their interpretations and reasoning.
Level 4 545	At level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilize their limited range of skills and can reason with some insight, in straightforward contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.
Level 3 482	At level 3, students can execute clearly described procedures, including those that require sequential decisions. Their interpretations are sufficiently sound to be a base for building a simple model or for selecting and applying simple problem-solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships. Their solutions reflect that they have engaged in basic interpretation and reasoning.
Level 2 420	At level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions to solve problems involving whole numbers. They are capable of making literal interpretations of the results.
Level 1 358	At level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are almost always obvious and follow immediately from the given stimuli.

NOTE: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into mathematics literacy levels according to their scores. Cut scores in the exhibit are rounded; exact cut scores are provided in appendix A. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Science Literacy

In PISA 2012, science literacy is defined as:

> *An individual's scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to draw evidence based conclusions about science-related issues; understanding of the characteristic features of science as a form of human knowledge and inquiry; awareness of how science and technology shape our material, intellectual, and cultural environments; and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen* (OECD 2013, p. 100).

Science literacy is reported both in terms of proficiency levels and scale scores (reported on a scale of 0–1,000). Exhibit 2 (see below) describes the six science literacy proficiency levels and their respective cut scores.

Exhibit 2. Description of PISA proficiency levels on science literacy scale: 2012

Proficiency level and lower cut score	Task descriptions
Level 6 708	At level 6, students can consistently identify, explain, and apply scientific knowledge and knowledge about science in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions. They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that center on personal, social, or global situations.
Level 5 633	At level 5, students can identify the scientific components of many complex life situations, apply both scientific concepts and knowledge about science to these situations, and can compare, select and evaluate appropriate scientific evidence for responding to life situations. Students at this level can use well-developed inquiry abilities, link knowledge appropriately, and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.
Level 4 559	At level 4, students can work effectively with situations and issues that may involve explicit phenomena requiring them to make inferences about the role of science or technology. They can select and integrate explanations from different disciplines of science or technology and link those explanations directly to aspects of life situations. Students at this level can reflect on their actions and they can communicate decisions using scientific knowledge and evidence.
Level 3 484	At level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.
Level 2 410	At level 2, students have adequate scientific knowledge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations. They are capable of direct reasoning and making literal interpretations of the results of scientific inquiry or technological problem solving.
Level 1 335	At level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and follow explicitly from given evidence.

NOTE: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into science literacy levels according to their scores. Cut scores in the exhibit are rounded; exact cut scores are provided in appendix A. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Reading Literacy

In PISA 2012, reading literacy is defined as:

> *Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society* (OECD 2013, p. 61).

Reading literacy is reported both in terms of proficiency levels and scale scores (reported on a scale of 0–1,000). Exhibit 3 (see following page) describes the seven reading literacy proficiency levels and their respective cut scores.

Exhibit 3. Description of PISA proficiency levels on reading literacy scale: 2012

Proficiency level and lower cut score	Task descriptions
Level 6 698	At level 6, tasks typically require the reader to make multiple inferences, comparisons, and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. Reflect and evaluate tasks may require the reader to hypothesize about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for access and retrieve tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.
Level 5 626	At level 5, tasks that involve retrieving information require the reader to locate and organize several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialized knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.
Level 4 553	At level 4, tasks that involve retrieving information require the reader to locate and organize several pieces of embedded information. Some tasks at this level require interpreting the meaning of nuances of language in a section of text by taking into account the text as a whole. Other interpretative tasks require understanding and applying categories in an unfamiliar context. Reflective tasks at this level require readers to use formal or public knowledge to hypothesize about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts whose content or form may be unfamiliar.
Level 3 480	At level 3, tasks require the reader to locate, and in some cases recognize the relationship between, several pieces of information that must meet multiple conditions. Interpretative tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship, or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorizing. Often the required information is not prominent or there is much competing information; or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons, and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.
Level 2 407	At level 2, some tasks require the reader to locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. Others require recognizing the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.
Level 1a 335	At level 1a, tasks require the reader to locate one or more independent pieces of explicitly stated information; to recognize the main theme or author’s purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically, the required information in the text is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.
Level 1b 262	At level 1b, tasks require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures, or familiar symbols. There is minimal competing information. In tasks requiring interpretation the reader may need to make simple connections between adjacent pieces of information.

NOTE: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into reading literacy levels according to their scores. Cut scores in the exhibit are rounded; exact cut scores are provided in appendix A. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012. NOTE: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into mathematics literacy levels according to their scores. Cut scores in the exhibit are rounded; exact cut scores are provided in appendix A. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Computer-Based Assessments

In 2012, computer-based assessments in mathematics and reading were offered as optional assessments for participating education systems. Thirty-two education systems, including the United States, chose to administer them. In these education systems, a subset of students who took the paper-based assessment also took an additional computer-based assessment. Although the paper-based assessment items and the computer-based assessment items were derived from the same frameworks, there was no overlap in the assessment items between the two assessment modes. The interactive nature of computer-based assessment allowed PISA to assess students in novel contexts that are not possible with a traditional paper-based format. For instance, the computer-based mathematics assessment was designed to measure the same mathematics content and processes as the paper-based assessment, but the computer-based environment provided the opportunity to include tasks requiring students to manipulate mathematical tools like statistical software, geometric construction, visualization utilities, and virtual measuring instruments (OECD 2013, pp. 43–44). And, while individuals use many of the same reading processes and skills when they are reading print or reading online, there are reading processes that are unique to an electronic environment, such as navigation across multiple sites without explicit direction or using web-based navigation tools such as drop-down menus (OECD 2013, p. 80). The computer-based reading assessment was designed to investigate students' proficiency in that context. For both mathematics and reading, the paper-based assessment and computer-based assessment were scaled separately. Therefore, scores on the paper-based assessment cannot be compared to scores on the computer-based assessment.

Reporting PISA 2012 Results

This report presents performance on PISA 2012 in mathematics, science, and reading literacy from a U.S. perspective. Results are presented for the 65 education systems, including the United States, that participated in PISA 2012 and for the three U.S. states—Connecticut, Florida, and Massachusetts—that participated as separate education systems. These three states opted to have separate samples of public-school schools and students included in PISA in order to obtain state-level results.

In this report, results are presented in terms of average scale scores and the percentage of 15-year-old students reaching selected proficiency levels, comparing the United States with other participating education systems. For proficiency levels, results are reported in terms of the percentage reaching level 5 or above and the percentage below level 2. Higher proficiency levels represent the knowledge, skills, and capabilities needed to perform tasks of greater complexity. At levels 5 and 6, students demonstrate higher level skills and may be referred to as "top performers" in the subject. Conversely, students performing below level 2 are below what the OECD calls "a baseline level of proficiency, at which students begin to demonstrate the…literacy competencies that will enable them to participate effectively and productively in life" (OECD 2010, p. 154).[2]

This report also presents U.S. trends over time in mathematics, science, and reading literacy and overall results for the computer-based mathematics and reading assessments. Results for the problem-solving and financial literacy assessments will be released in 2014.

In reporting PISA results, the OECD differentiates between OECD member countries, of which there are 34, and all other participating education systems, some of which are countries and some

[2] Percentages of students at each proficiency level may be found at http://nces.ed.gov/surveys/pisa/pisa2012/.

of which are subnational entities. In the OECD's PISA reports, OECD member countries and other participating education systems are reported in the tables and figures in the main body of the report, along with the average for the OECD countries (the average of all OECD member country averages with each country weighted equally), and are discussed in the accompanying text. Also, for some participating education systems, results for subnational entities—including, in 2012, the three U.S. states—are reported in appendixes of the OECD PISA reports but are not discussed in the report text.

To facilitate readers moving between the OECD and U.S. national PISA reports, this report's tables and figures follow the OECD convention of placing OECD member countries and all other participating education systems in the main part of the tables and figures. These are all referred to as education systems in this report, and there are 65 altogether. The three U.S. states that participated in PISA 2012 are presented in a separate part of the tables and figures; results for the states are discussed in the text but are not included in counts of education systems performing above, below, or not measurably different from the United States.

This report is merely a first look at the PISA 2012 results and is by no means comprehensive. For more PISA 2012 results, visit the National Center for Education Statistics PISA website at http://nces.ed.gov/surveys/pisa/pisa2012/. The website includes more results from the mathematics, reading, science, and computer-based assessments, including results for various subgroups (e.g., by gender, race/ethnicity), in mathematics subscales, and on trends in performance, and more detailed results for the three U.S. states that participated in 2012.

All statistically significant differences described in this report are at the .05 level. Differences that are statistically significant are discussed using comparative terms such as "higher" and "lower." Differences that are not statistically significant are either not discussed or referred to as "not measurably different." In almost all instances, the tests for significance used were standard *t* tests (see appendix A for additional details on interpreting statistical significance). No adjustments were made for multiple comparisons.

Selected Findings

U.S. Performance in Mathematics Literacy

- Percentages of top performing 15-year-old students (those scoring at level 5 or above) in mathematics literacy ranged from 55 percent in Shanghai-China to nearly 0 percent in Colombia and Argentina. In the United States, 9 percent of 15-year-old students scored at proficiency level 5 or above, which was lower than the OECD average of 13 percent. The U.S. percentage was lower than 27 education systems, higher than 22 education systems, and not measurably different than 13 education systems. The percentage of top performers in mathematics in the United States overall (9 percent) was higher than the state of Florida (6 percent), but lower than Massachusetts (19 percent) and Connecticut (16 percent) (figure 1).
- In mathematics literacy, the percentage of 15-year-old students performing below level 2, which is considered a baseline of proficiency by the OECD, ranged from 4 percent in Shanghai-China to 76 percent in Indonesia. In the United States, 26 percent of 15-year-old students scored below level 2, which was higher than the OECD average of 23 percent. The U.S. percentage was higher than 29 education systems, lower than 26 education systems, and not measurably different than 9 education systems. The percentage of low performers in mathematics in the United States overall (26 percent) was higher than the states of Connecticut (21 percent) and Massachusetts (18 percent), but not measurably different than Florida (30 percent) (figure 1).
- Average scores in mathematics literacy ranged from 613 in Shanghai-China to 368 in Peru. The U.S. average score was 481, which was lower than the OECD average of 494. The U.S. average was lower than 29 education systems, higher than 26 education systems, and not measurably different than 9 education systems. The U.S. average was lower than the states of Massachusetts (514) and Connecticut (506), but higher than Florida (467) (table 1).

U.S. Performance in Science Literacy

- Percentages of top-performing 15-year-old students (those scoring at level 5 or above) in science literacy ranged from 27 percent in Shanghai-China and 23 percent in Singapore to nearly 0 percent in eight education systems. In the United States, 7 percent of 15-year-old students scored at proficiency level 5 or above, which was not measurably different from the OECD average of 8 percent. The U.S. percentage was lower than 17 education systems, higher than 27 education systems, and not measurably different than 15 education systems. The percentage of top performers in science in the United States overall (7 percent) was lower than the states of Massachusetts (14 percent) and Connecticut (13 percent), but not measurably different than Florida (5 percent) (figure 2).
- In science literacy, the percentage of 15-year-old students performing below level 2, which is considered a baseline of proficiency by the OECD, ranged from 3 percent in Shanghai-China and 5 percent in Estonia to 67 percent in Indonesia and 68 percent in Peru. In the United States, 18 percent of U.S. 15-year-old students scored below level 2, which was not measurably different from the OECD average of 18 percent. The U.S. percentage was higher than 21 education systems, lower than 29 education systems, and not measurably different than 14 education systems. The percentage of low performers in science in the United States overall (18 percent) was higher than the states of Connecticut (13 percent) and Massachusetts (11 percent), but not measurably different than Florida (21 percent) (figure 2).

- Average scores in science literacy ranged from 580 in Shanghai-China to 373 in Peru. The U.S. average score was 497, which was not measurably different from the OECD average of 501. The U.S. average was lower than 22 education systems, higher than 29 education systems, and not measurably different than 13 education systems. The U.S. average was lower than the states of Massachusetts (527) and Connecticut (521), but not measurably different than Florida (485) (table 2).

U.S. Performance in Reading Literacy

- Percentages of top performing 15-year-old students (those scoring at level 5 or above) in reading literacy ranged from 25 percent in Shanghai-China and 21 percent in Singapore to nearly 0 percent in 3 education systems. In the United States, 8 percent of U.S. 15-year-old students scored at proficiency level 5 or above, which was not measurably different from the OECD average of 8 percent. The U.S. percentage was lower than 14 education systems, higher than 33 education systems, and not measurably different than 12 education systems. The percentage of top performers in reading in the United States overall (8 percent) was higher than the state of Florida (6 percent), but lower than Massachusetts (16 percent) and Connecticut (15 percent) (figure 3).
- In reading literacy, the percentage of 15-year-old students performing below level 2, which is considered a baseline of proficiency by the OECD, ranged from 3 percent in Shanghai-China to 60 percent in Peru. In the United States, 17 percent of U.S. 15-year-old students scored below level 2, which was not measurably different from the OECD average of 18 percent. The U.S. percentage was higher than 14 education systems, lower than 33 education systems, and not measurably different than 17 education systems. The percentage of low performers in reading in the United States overall (17 percent) was higher than the state of Massachusetts (11 percent), but not measurably different than Connecticut (13 percent) and Florida (17 percent) (figure 3).
- Average scores in reading literacy ranged from 570 in Shanghai-China to 384 in Peru. The U.S. average score was 498, which was not measurably different from the OECD average of 496. The U.S. average was lower than 19 education systems, higher than 34 education systems, and not measurably different than 11 education systems. The U.S. average was lower than the U.S. states Massachusetts (527) and Connecticut (521), but not measurably different than Florida (492) (table 3).

Eighteen education systems had higher average scores than the United States in *all three subjects*. The 18 education systems are: Australia, Canada, Chinese Taipei, Estonia, Finland, Germany, Hong Kong-China, Ireland, Japan, Liechtenstein, Macao-China, Netherlands, New Zealand, Poland, Republic of Korea, Shanghai-China, Singapore, and Switzerland. The U.S. states Massachusetts and Connecticut also had higher average scores than the United States in *all three subjects* (tables 1, 2, and 3).

U.S. Performance Over Time

- The U.S. average mathematics, science, and reading literacy scores in 2012 were not measurably different from average scores in previous PISA assessment years with which comparisons can be made (2003, 2006, and 2009 for mathematics; 2006 and 2009 for science; and 2000, 2003, and 2009 for reading) (table 4).

U.S. Performance on Computer-Based Assessments

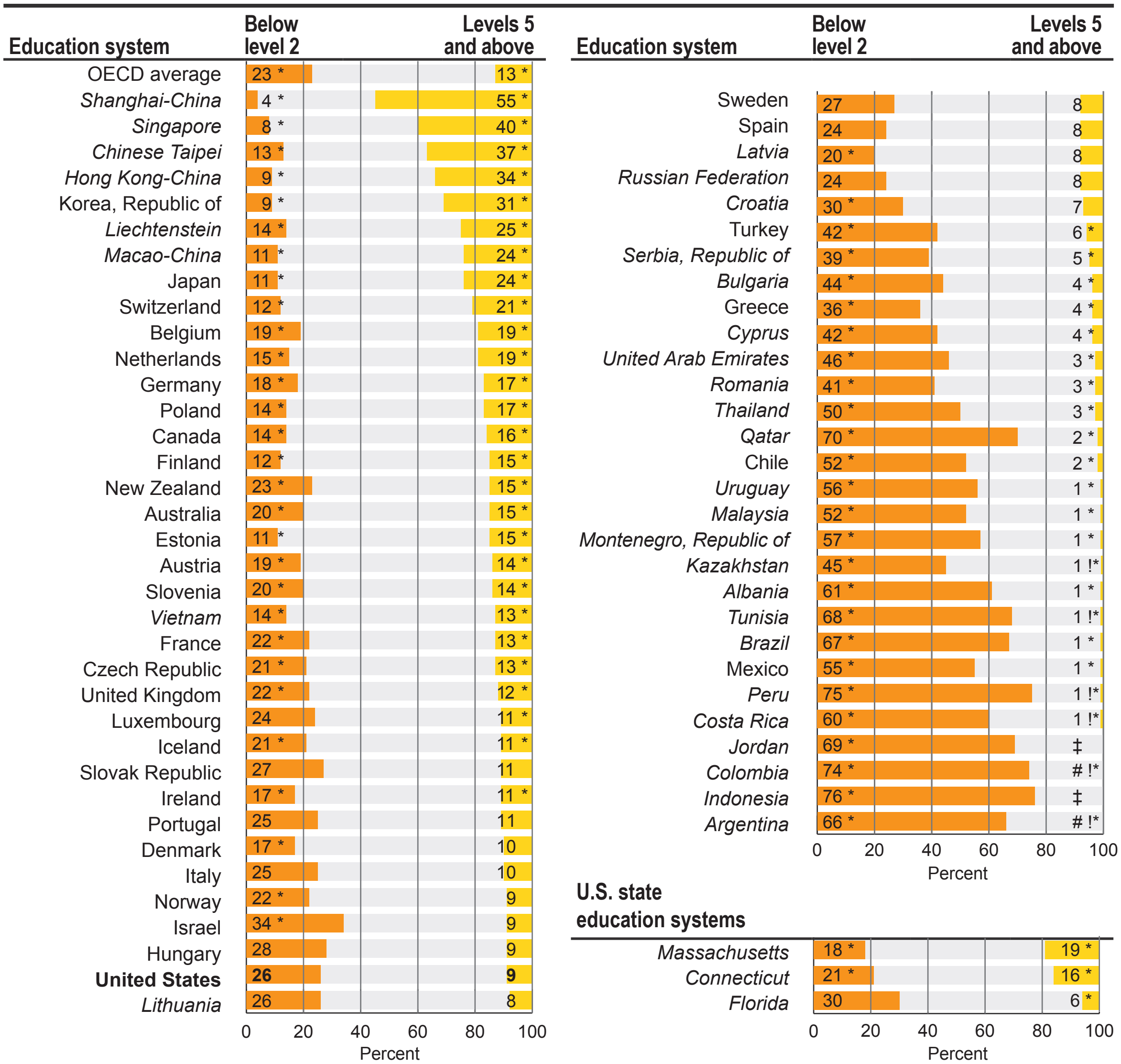
- On the computer-based mathematics literacy assessment (administered in 32 education systems), average scores ranged from 566 in Singapore and 562 in Shanghai-China to 397 in Colombia. U.S. 15-year-old students had an average score of 498, which was not measurably different from the OECD average of 497. Twelve education systems had higher average scores, 8 had lower average scores, and 11 had average scores that were not measurably different than the United States (table 5).

- On the computer-based reading literacy assessment (administered in 32 education systems), average scores ranged from 567 in Singapore to 396 in Colombia. U.S. 15-year-old students had an average score of 511, which was higher than the OECD average of 497. Seven education systems had higher average scores, 17 had lower average scores, and 7 had average scores that were not measurably different than the United States (table 6).

This page intentionally left blank

Figures and Tables

Figure 1. Percentage of 15-year-old students performing at PISA mathematics literacy proficiency levels 5 and above and below level 2, by education system: 2012



Below level 2
Levels 5 and above
Rounds to zero.
! Interpret with caution. Estimate is unstable due to high coefficient of variation.
‡ Reporting standards not met.
* $p < .05$. Significantly different from the U.S. percentage at the .05 level of significance.
NOTE: Education systems are ordered by 2012 percentages of 15-year-olds in levels 5 and above. To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into mathematics proficiency levels according to their scores. Cut scores for each proficiency level can be found in table A-1 in appendix A. The OECD average is the average of the national percentages of the OECD member countries, with each country weighted equally. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table M1b available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

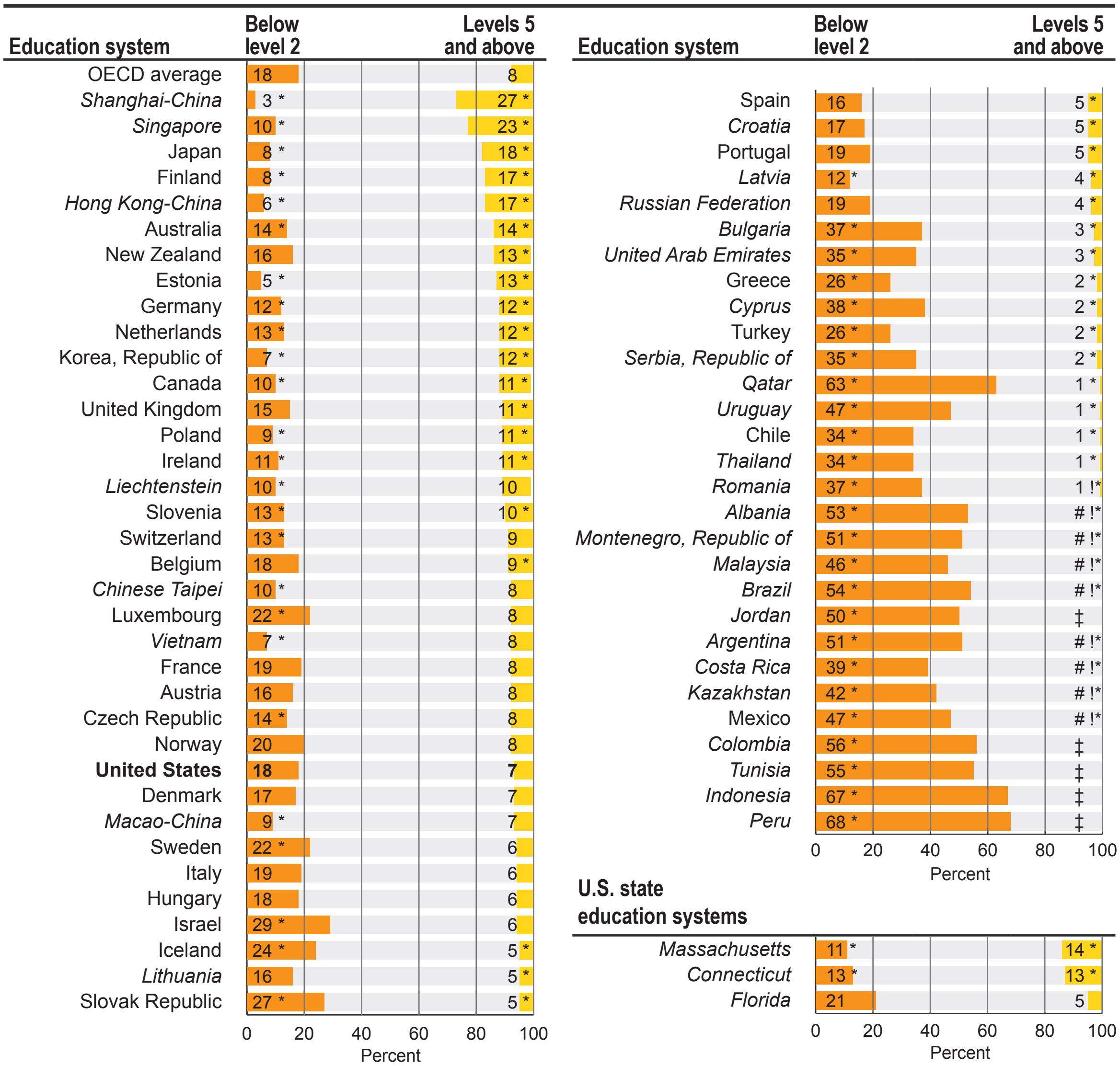
Table 1. Average scores of 15-year-old students on PISA mathematics literacy scale, by education system: 2012

Education system	Average score	
OECD average	494	▲
Shanghai-China	613	▲
Singapore	573	▲
Hong Kong-China	561	▲
Chinese Taipei	560	▲
Korea, Republic of	554	▲
Macao-China	538	▲
Japan	536	▲
Liechtenstein	535	▲
Switzerland	531	▲
Netherlands	523	▲
Estonia	521	▲
Finland	519	▲
Canada	518	▲
Poland	518	▲
Belgium	515	▲
Germany	514	▲
Vietnam	511	▲
Austria	506	▲
Australia	504	▲
Ireland	501	▲
Slovenia	501	▲
Denmark	500	▲
New Zealand	500	▲
Czech Republic	499	▲
France	495	▲
United Kingdom	494	▲
Iceland	493	▲
Latvia	491	▲
Luxembourg	490	▲
Norway	489	
Portugal	487	
Italy	485	
Spain	484	
Russian Federation	482	
Slovak Republic	482	
United States	**481**	
Lithuania	479	
Sweden	478	
Hungary	477	
Croatia	471	▼
Israel	466	▼
Greece	453	▼
Serbia, Republic of	449	▼
Turkey	448	▼
Romania	445	▼
Cyprus	440	▼
Bulgaria	439	▼
United Arab Emirates	434	▼
Kazakhstan	432	▼
Thailand	427	▼
Chile	423	▼
Malaysia	421	▼
Mexico	413	▼
Montenegro, Republic of	410	▼
Uruguay	409	▼
Costa Rica	407	▼
Albania	394	▼
Brazil	391	▼
Argentina	388	▼
Tunisia	388	▼
Jordan	386	▼
Colombia	376	▼
Qatar	376	▼
Indonesia	375	▼
Peru	368	▼

U.S. state education systems		
Massachusetts	514	▲
Connecticut	506	▲
Florida	467	▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
NOTE: Education systems are ordered by 2012 average score. The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table M4 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024 .
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Figure 2. Percentage of 15-year-old students performing at PISA science literacy proficiency levels 5 and above and below level 2, by education system: 2012



Below level 2
Levels 5 and above
Rounds to zero.
! Interpret with caution. Estimate is unstable due to high coefficient of variation.
‡ Reporting standards not met.
* $p < .05$. Significantly different from the U.S. percentage at the .05 level of significance.
NOTE: Education systems are ordered by 2012 percentages of 15-year-olds in levels 5 and above. To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into science proficiency levels according to their scores. Cut scores for each proficiency level can be found in table A-1 in appendix A. The OECD average is the average of the national percentages of the OECD member countries, with each country weighted equally. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table S1b available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

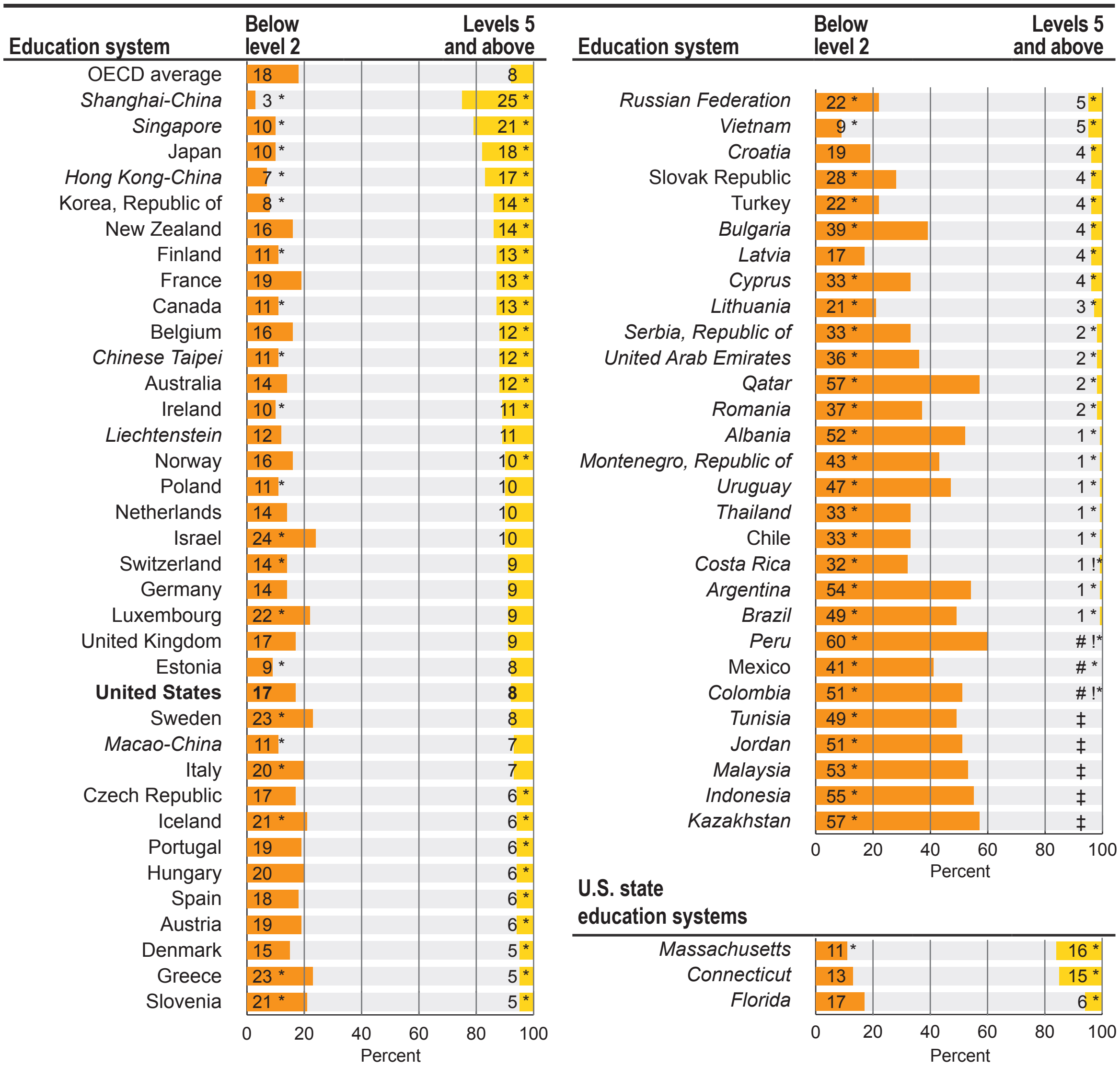
Table 2. Average scores of 15-year-old students on PISA science literacy scale, by education system: 2012

Education system	Average score
OECD average	501
Shanghai-China	580 ▲
Hong Kong-China	555 ▲
Singapore	551 ▲
Japan	547 ▲
Finland	545 ▲
Estonia	541 ▲
Korea, Republic of	538 ▲
Vietnam	528 ▲
Poland	526 ▲
Canada	525 ▲
Liechtenstein	525 ▲
Germany	524 ▲
Chinese Taipei	523 ▲
Netherlands	522 ▲
Ireland	522 ▲
Australia	521 ▲
Macao-China	521 ▲
New Zealand	516 ▲
Switzerland	515 ▲
Slovenia	514 ▲
United Kingdom	514 ▲
Czech Republic	508 ▲
Austria	506
Belgium	505
Latvia	502
France	499
Denmark	498
United States	**497**
Spain	496
Lithuania	496
Norway	495
Hungary	494
Italy	494
Croatia	491
Luxembourg	491
Portugal	489
Russian Federation	486 ▼
Sweden	485 ▼
Iceland	478 ▼
Slovak Republic	471 ▼
Israel	470 ▼
Greece	467 ▼
Turkey	463 ▼
United Arab Emirates	448 ▼
Bulgaria	446 ▼
Chile	445 ▼
Serbia, Republic of	445 ▼
Thailand	444 ▼
Romania	439 ▼
Cyprus	438 ▼
Costa Rica	429 ▼
Kazakhstan	425 ▼
Malaysia	420 ▼
Uruguay	416 ▼
Mexico	415 ▼
Montenegro, Republic of	410 ▼
Jordan	409 ▼
Argentina	406 ▼
Brazil	405 ▼
Colombia	399 ▼
Tunisia	398 ▼
Albania	397 ▼
Qatar	384 ▼
Indonesia	382 ▼
Peru	373 ▼

U.S. state education systems	
Massachusetts	527 ▲
Connecticut	521 ▲
Florida	485

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
NOTE: Education systems are ordered by 2012 average score. The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table S2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Figure 3. Percentage of 15-year-old students performing at PISA reading literacy proficiency levels 5 and above and below level 2, by education system: 2012



Below level 2
Levels 5 and above
Rounds to zero.
! Interpret with caution. Estimate is unstable due to high coefficient of variation.
‡ Reporting standards not met.
* *p* < .05. Significantly different from the U.S. percentage at the .05 level of significance.
NOTE: Education systems are ordered by 2012 percentages of 15-year-olds in levels 5 and above. To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into reading proficiency levels according to their scores. Cut scores for each proficiency level can be found in table A-1 in appendix A. The OECD average is the average of the national percentages of the OECD member countries, with each country weighted equally. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table R1b available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024. .
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Table 3. Average scores of 15-year-old students on PISA reading literacy scale, by education system: 2012

Education system	Average score	
OECD average	496	
Shanghai-China	570	▲
Hong Kong-China	545	▲
Singapore	542	▲
Japan	538	▲
Korea, Republic of	536	▲
Finland	524	▲
Ireland	523	▲
Chinese Taipei	523	▲
Canada	523	▲
Poland	518	▲
Estonia	516	▲
Liechtenstein	516	▲
New Zealand	512	▲
Australia	512	▲
Netherlands	511	▲
Belgium	509	▲
Switzerland	509	▲
Macao-China	509	▲
Vietnam	508	
Germany	508	▲
France	505	
Norway	504	
United Kingdom	499	
United States	**498**	
Denmark	496	
Czech Republic	493	
Italy	490	
Austria	490	
Latvia	489	▼
Hungary	488	
Spain	488	▼
Luxembourg	488	▼
Portugal	488	
Israel	486	
Croatia	485	▼
Sweden	483	▼
Iceland	483	▼
Slovenia	481	▼
Lithuania	477	▼
Greece	477	▼
Turkey	475	▼
Russian Federation	475	▼
Slovak Republic	463	▼
Cyprus	449	▼
Serbia, Republic of	446	▼
United Arab Emirates	442	▼
Chile	441	▼
Thailand	441	▼
Costa Rica	441	▼
Romania	438	▼
Bulgaria	436	▼
Mexico	424	▼
Montenegro, Republic of	422	▼
Uruguay	411	▼
Brazil	410	▼
Tunisia	404	▼
Colombia	403	▼
Jordan	399	▼
Malaysia	398	▼
Indonesia	396	▼
Argentina	396	▼
Albania	394	▼
Kazakhstan	393	▼
Qatar	388	▼
Peru	384	▼
U.S. state education systems		
Massachusetts	527	▲
Connecticut	521	▲
Florida	492	

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
NOTE: Education systems are ordered by 2012 average score. The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only. The standard errors of the estimates are shown in table R2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Table 4. Average scores and changes in average scores of U.S. 15-year-old students on PISA mathematics, science, and literacy scales: 2000, 2003, 2006, 2009, and 2012

	Average score					Change in average score			
Subject	2000	2003	2006	2009	2012	2000–2012	2003–2012	2006–2012	2009–2012
Mathematics literacy	†	483	474	487	481	†	○	○	○
Science literacy	†	†	489	502	497	†	†	○	○
Reading literacy	504	495	—	500	498	○	○	—	○

○ Average score in 2012 is not measurably different than average score in comparison year.
— Not available. PISA 2006 reading literacy results are not reported for the United States because of an error in printing the test booklets and comparisons are not possible.
† Not applicable. Although mathematics was assessed in 2000 and science was assessed in 2000 and 2003, because the mathematics framework was revised for PISA 2003 and the science framework was revised for 2006, it is possible to look at changes in mathematics only from 2003 forward and in science only from 2006 forward.
NOTE: All average scores reported as higher or lower than the comparison year are different at the .05 level of statistical significance. The standard errors of the estimates are shown in table T1 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, 2006, 2009, 2012.

Table 5. Average scores of 15-year-old students on PISA computer-based mathematics literacy scale, by education system: 2012

Education system	Average score		Education system	Average score	
OECD average	497				
Singapore	566	▲	Norway	498	
Shanghai-China	562	▲	Slovak Republic	497	
Korea, Republic of	553	▲	Denmark	496	
Hong Kong-China	550	▲	Ireland	493	
Macao-China	543	▲	Sweden	490	
Japan	539	▲	*Russian Federation*	489	
Chinese Taipei	537	▲	Poland	489	
Canada	523	▲	Portugal	489	
Estonia	516	▲	Slovenia	487	▼
Belgium	511	▲	Spain	475	▼
Germany	509	▲	Hungary	470	▼
France	508		Israel	447	▼
Australia	508	▲	*United Arab Emirates*	434	▼
Austria	507		Chile	432	▼
Italy	499		*Brazil*	421	▼
United States	**498**		*Colombia*	397	▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
NOTE: The computer-based mathematics literacy assessment was an optional assessment for education systems in 2012. Education systems are ordered by 2012 average score. The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. Italics indicate non-OECD countries and education systems. The standard errors of the estimates are shown in table CM2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Table 6. Average scores of 15-year-old students on PISA computer-based reading literacy scale, by education system: 2012

Education system	Average score	Education system	Average score
OECD average	497 ▼		
Singapore	567 ▲	Sweden	498 ▼
Korea, Republic of	555 ▲	Denmark	495 ▼
Hong Kong-China	550 ▲	Germany	494 ▼
Japan	545 ▲	Portugal	486 ▼
Canada	532 ▲	Austria	480 ▼
Shanghai-China	531 ▲	Poland	477 ▼
Estonia	523 ▲	Slovak Republic	474 ▼
Australia	521	Slovenia	471 ▼
Ireland	520	Spain	466 ▼
Chinese Taipei	519	*Russian Federation*	466 ▼
Macao-China	515	Israel	461 ▼
United States	**511**	Chile	452 ▼
France	511	Hungary	450 ▼
Italy	504	*Brazil*	436 ▼
Belgium	502	*United Arab Emirates*	407 ▼
Norway	500 ▼	*Colombia*	396 ▼

▲ Average score is higher than U.S. average score.
▼ Average score is lower than U.S. average score.
NOTE: The computer-based reading literacy assessment was an optional assessment for education systems in 2012. Education systems are ordered by 2012 average score. The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. All average scores reported as higher or lower than the U.S. average score are different at the .05 level of statistical significance. Italics indicate non-OECD countries and education systems. The standard errors of the estimates are shown in table CM2 available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

References

Australian Council for Educational Research (ACER). (2011). *PISA 2012 Main Survey Test Administrator's Manual (Core PISA)*. Camberwell: Author.

Judkins, D.R. (1990). Fay's Method for Variance Estimation. *Journal of Official Statistics 6*(3): 223-239.

Kastberg, D., Roey, S., Lemanski, N., Chan, J.Y., and Murray, G. (forthcoming). *Technical Report and User Guide for the Program for International Student Assessment (PISA).* U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office.

Krenzke, T., Roey, S., Dohrmann, S., Mohadjer, L., Haung, W., Kaufman, S., and Seastrom, M. (2006). Tactics for Reducing the Risk of Disclosure Using the NCES DataSwap Software. *Proceedings of the Survey Research Methods Section 2006*. Alexandria, VA: American Statistical Association.

Organization for Economic Cooperation and Development (OECD). (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010). *PISA 2009 Results: What Students Know and Can Do—Student Performance in Reading, Mathematics and Science (Volume I).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2013). *PISA 2012 Assessment Framework—Key Competencies in Reading, Mathematics and Science*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (forthcoming). *PISA 2012 Technical Report*. Paris: Author.

Appendix A: Methodology and Technical Notes

This appendix describes features of the Program for International Student Assessment (PISA) 2012 methodology, including sample design, test design, and scoring, with a particular focus on the U.S. implementation. For further details about the assessment and any of the topics discussed here, see the Organization for Economic Cooperation and Development's (OECD) *PISA 2012 Technical Report* (forthcoming).

International Requirements for Sampling, Data Collection, and Response Rates

OECD required all participating education systems to adhere to the PISA 2012 technical standards (OECD forthcoming), which provided detailed information about the target population, sampling, response rates, translation, assessment administration, and data submission. According to the standards, the international desired population in each education system consisted of 15-year-olds attending both publicly and privately controlled schools in grade 7 and higher. To provide valid estimates of student achievement and characteristics, the sample of PISA students had to be selected in a way that represented the full population of 15-year-old students in each education system. The sample design for PISA 2012 was a stratified systematic sample, with sampling probabilities proportional to the estimated number of 15-year-old students in the school based on grade enrollments. Samples were drawn using a two-stage sampling process. The first stage was a sample of schools, and the second stage was a sample of students within schools. The PISA international contractors responsible for the design and implementation of PISA internationally (hereafter referred to as the PISA consortium) drew the sample of schools for each education system.

A minimum of 4,500 students from a minimum of 150 schools was required in each country.[1] Following the PISA consortium guidelines, replacement schools were identified at the same time the PISA sample was selected by assigning the two schools neighboring the sampled school in the frame as replacements. The international guidelines specified that within schools, a sample of 35 students was to be selected in an equal probability sample unless fewer than 35 students age 15 were available (in which case all 15-year-old students were selected).

Each education system collected its own data, following international guidelines and specifications. The technical standards required that students in the sample be 15 years and 3 months to 16 years and 2 months at the beginning of the testing period. The maximum length of the testing period was 42 days. Most education systems conducted testing from March through August 2012.[2]

The school response-rate target was 85 percent for all education systems. This target applies in aggregate, not to each individual school. A minimum of 65 percent of schools from the original sample of schools was required to participate for an education system's data to be included in the international database. Education systems were allowed to use replacement

[1] PISA also includes education systems that are not countries, such as Hong Kong and Shanghai in China. Non-national entities were required to sample a minimum of 1,500 students from at least 50 schools. In the United States, three states (Connecticut, Florida, and Massachusetts) provided state-level samples in addition to the schools for the national sample in order to obtain state-level PISA estimates.

[2] The United States and the United Kingdom were given permission to move the testing dates to September through November in an effort to improve response rates. The range of eligible birth dates was adjusted so that the mean age remained the same (i.e., 15 years and 3 months to 16 years and 2 months at the beginning of the testing period). In 2003, the United States conducted PISA in the spring and fall and found no significant difference in student performance between the two time points.

schools (selected during the sampling process) to increase the response rate once the 65 percent benchmark had been reached. Replacement students within a school were not allowed.

The technical standards also required a minimum participation rate of 80 percent of sampled students from schools (sampled and replacement) within each education system. Follow-up sessions were required in schools where too few students participated in the originally scheduled test sessions to ensure a high overall student response rate. A student was considered to be a participant if he or she participated in the first testing session or a follow-up or makeup testing session. Data from education systems not meeting this requirement could be excluded from international reports. See appendix B for final response rates by education system.

PISA 2012 is designed to be as inclusive as possible. The guidelines allowed schools to be excluded for approved reasons (for example, schools in remote regions, very small schools, or special education schools). Schools used the following international guidelines on student exclusions:

- **Students with functional disabilities.** These were students with a moderate to severe permanent physical disability such that they cannot perform in the PISA testing environment.
- **Students with intellectual disabilities.** These were students with a mental or emotional disability and who have been tested as cognitively delayed or who are considered in the professional opinion of qualified staff to be cognitively delayed such that they cannot perform in the PISA testing environment.
- **Students with insufficient language experience.** These were students who meet the three criteria of not being native speakers in the assessment language, having limited proficiency in the assessment language, and having less than 1 year of instruction in the assessment language.

Overall estimated exclusions (including both school and student exclusions) were to be under 5 percent of the PISA target population.

Sampling and Data Collection in the United States

The PISA 2012 school sample was drawn for the United States by the PISA consortium. The U.S. PISA sample was stratified into eight explicit groups based on control of school (public or private) and region of the country (Northeast, Central, West, Southeast).[3] Within each stratum, the frame was sorted for sampling by five categorical stratification variables: grade range of the school (five categories); type of location relative to populous areas (city, suburb, town, rural);[4] combined percentage of Black, Hispanic, Asian, Native Hawaiian/Pacific Islander, and American Indian/Alaska Native students (above or below 15 percent); gender (mostly female [percent female >= 95 percent], mostly male [percent female < 5 percent]; and other); and state. The same frame and characteristics were used for the state samples.

For the U.S. national sample, within each school, 50 students aged 15 were randomly sampled. The United States increased its national sample from the international standard of 35 to 50 in order

[3] The Northeast region consists of Connecticut, Delaware, the District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Central region consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Wisconsin, and South Dakota. The West region consists of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming. The Southeast region consists of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

[4] These types are defined as follows: (1) "city" is a territory inside an urbanized area with a core population of 50,000 or more and inside a principal city; (2) "suburb" is a territory inside an urbanized area with a core population of 50,000 or more and outside a principal city; (3) "town" is a territory inside an urban cluster with a core population between 25,000 and 50,000; and (4) "rural" is a territory not in an urbanized area or urban cluster.

to reach the required number of students and in order to administer the optional financial literacy assessment. Connecticut, Florida, and Massachusetts participated in PISA 2012 with separate state samples drawn by the PISA consortium. The state samples are not part of the main sample. In each of the three state samples, 42 students aged 15 were randomly sampled within each school. If fewer than 50 age-eligible students (in schools in the national sample) or fewer than 42 age-eligible students (in schools in the state samples) were enrolled, all 15-year-old students in a school were selected. Thus, in each school, each age-eligible student had an equal probability of being selected. Sampled students were born between July 1, 1996, and June 30, 1997 (hereafter the sampled students are referred to as "15-year-olds" or "15-year-old students").

In the national-sample schools, of the 50 students, 42 took the paper-based mathematics, science, and reading literacy assessments, 20 of which were subsampled to also take the computer-based assessment, and 8 took the financial literacy assessment. In the state sample schools, all sampled students took only the paper-based mathematics, science, and reading assessments. The technical standard for the maximum length of the testing period was 42 days, but the United States requested and was granted permission to expand the testing window to 60 days (from October 2, 2012, to November 30, 2012) to accommodate school requests.

The U.S. PISA 2012 national school sample consisted of 240 schools.[5] This number was increased from the international minimum requirement of 150 to offset school nonresponse and reduce design effects. Schools were selected with probability proportionate to the school's estimated enrollment of 15-year-olds. The data for public schools were from the 2008–09 Common Core of Data and the data for private schools were from the 2009–10 Private School Universe Survey. Any school containing at least one 7th- through 12th-grade class was included in the school sampling frame. Participating schools provided a list of 15-year-old students (typically in August or September 2012) from which the sample was drawn using sampling software provided by the international contractor.

In addition to the international response rate standards described in the prior section, the U.S. sample had to meet the statistical standards of the National Center for Education Statistics (NCES) of the U.S. Department of Education. For assessments, NCES requires that the response rate should be at least 80 percent for schools and at least 85 percent for students.

Test Development

The 2012 assessment instruments were developed by international experts and PISA consortium test developers and included items submitted by participating education systems. Items were reviewed by representatives of each country for possible bias and relevance to PISA's goals and the PISA subject-matter expert groups. All participating education systems field-tested the assessment items in spring 2011.

The final paper-based assessment consisted of 85 mathematics items, 44 reading items, 53 science items, and 40 financial literacy items allocated to 17 test booklets (in education systems that did not administer the optional financial literacy assessment there were 13 test booklets). Each booklet was made up of four test clusters. Altogether there were seven mathematics clusters, three reading clusters, three science clusters, and two financial literacy clusters. The mathematics, science, and reading clusters were allocated in a rotated design to 13 booklets. The financial literacy clusters

[5] The state samples consisted of 54, 55, and 54 schools for Connecticut, Florida, and Massachusetts, respectively. As with the PISA national sample, these numbers were increased from the international minimum of 50 schools for subnational entities to offset school nonresponse and ineligibility.

in conjunction with mathematics and reading clusters were allocated in a rotated design to four booklets. The average number of items per cluster was 12 items for mathematics, 15 items for reading, 18 items for science, and 20 items for financial literacy. Each cluster was designed to average 30 minutes of test material. Each student took one booklet, with about 2 hours' worth of testing material. Approximately half of the items were multiple-choice, about 20 percent were closed or short response types (for which students wrote an answer that was simply either correct or incorrect), and about 30 percent were open constructed responses (for which students wrote answers that were graded by trained scorers using an international scoring guide). In PISA 2012, with the exception of students participating in the financial literacy assessment, every student answered mathematics items. Not all students answered reading, science items, and\or financial literacy items.

A subset of students who took the paper-based assessment also took a 40-minute computer-based assessment. In the United States, the computer-based assessment consisted of problem solving and the optional computer-based assessment of mathematics and reading. The computer-based assessment consisted of 168 problem-solving items, 164 mathematics items, and 144 reading items allocated to 24 forms. Each form was made up of two clusters that together contained 18 to 22 items. Altogether there were four clusters of problem solving, four clusters of mathematics, and two clusters of reading. The problem-solving, mathematics, and reading clusters were allocated in a rotated design to the 24 forms. Each cluster was designed to average 20 minutes of test material. (Not all education systems participated in the computer-based assessment and some education systems only administered the computer-based problem-solving assessment. Education systems that administered only the problem-solving assessment followed a different rotation design.)

In addition to the cognitive assessment, students also completed a 30-minute questionnaire designed to provide information about their backgrounds, attitudes, and experiences in school. Principals in schools where PISA was administered also completed a 30-minute questionnaire about their schools.

Translation and Adaptation

Source versions of all instruments (assessment booklets, computer-based assessment forms, questionnaires, and manuals) were prepared in English and French and translated into the primary language or languages of instruction in each education system. The PISA consortium recommended that education systems prepare and consolidate independent translations from both source versions and provided precise translation guidelines that included a description of the features each item was measuring and statistical analysis from the field trial. In cases for which one source language was used, independent translations were required and discrepancies reconciled. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in nations such as the United States that use English as the primary language of instruction. For example, words such as "lift" might be adapted to "elevator" for the United States. The PISA consortium verified the national adaptation of all instruments. Electronic copies of printed materials were sent to the PISA consortium for a final visual check prior to data collection.

Test Administration and Quality Assurance

The PISA consortium emphasized the use of standardized procedures in all education systems. Each education system collected its own data, based on a manual provided by the PISA consortium (ACER 2011) that explained the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators to use in testing sessions. Test administration

in the United States was conducted by professional staff trained in accordance with the international guidelines. Students were allowed to use calculators, and U.S. students were provided calculators.

In a sample of schools in each education system, a PISA Quality Monitor (PQM) who was engaged by the PISA consortium observed test administrations. The sample schools were selected jointly by the PISA consortium and the PQM. In the United States, there were two PQMs who each observed seven schools from the national and state samples. The PQM's primary responsibility was to document the extent to which testing procedures in schools were implemented in accordance with test administration procedures. The PQM's observations in U.S. schools indicated that international procedures for data collection were applied consistently.

Weighting

The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimates. Adjusted survey weights adjust for the probabilities of selection for individual schools and students, for school or student nonresponse, and for errors in estimating the size of the school or the number of 15-year-olds in the school at the time of sampling. Survey weighting for all education systems participating in PISA 2012 was coordinated by Westat, as part of the PISA consortium.

The school base weight was defined as the reciprocal of the school's probability of selection multiplied by the number of eligible students in the school. (For replacement schools, the school base weight was set equal to the original school it replaced.) The student base weight was given as the reciprocal of the probability of selection for each selected student from within a school.

The product of these base weights was then adjusted for school and student nonresponse. The school nonresponse adjustment was done individually for each education system by cross-classifying the explicit and implicit stratification variables defined as part of the sample design. Usually about 10 to 15 such cells were formed per education system.

The student nonresponse adjustment was done within cells based first on their school nonresponse cell and their explicit stratum; within that, grade and sex were used when possible. All PISA analyses were conducted using these adjusted sampling weights. For more information on the nonresponse adjustments, see OECD's *PISA 2012 Technical Report* (forthcoming).

Scaling of Student Test Data

Each test booklet or computerized version had a different subset of items. The fact that each student completed only a subset of items means that classical test scores, such as the percentage correct, are not accurate measures of student performance. Instead, scaling techniques were used to establish a common scale for all students. For PISA 2012, item response theory (IRT) was used to estimate average scores for mathematics, science, and reading literacy for each education system, as well as for three mathematics process and four mathematics content scales. For education systems participating in the financial literacy assessment and the computer-based assessment, these assessments will be scaled separately and assigned separate scores.

IRT identifies patterns of response and uses statistical models to predict the probability of answering an item correctly as a function of the students' proficiency in answering other questions. With this method, the performance of a sample of students in a subject area or subarea can be summarized on a simple scale or series of scales, even when students are administered different items.

Scores for students are estimated as plausible values because each student completed only a subset of items. Five plausible values were estimated for each student for each scale. These values represent the distribution of potential scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the PISA reading, mathematics, and science literacy scales are based on plausible values.

Proficiency Levels

In addition to a range of scale scores as the basic form of measurement, PISA describes student proficiency in terms of levels. Higher levels represent the knowledge, skills, and capabilities needed to perform tasks of increasing complexity. PISA results are reported in terms of percentages of the student population at each of the predefined levels.

To determine the performance levels and cut scores on the literacy scales, IRT techniques were used. With IRT techniques, it is possible to simultaneously estimate the ability of all students taking the PISA assessment, as well as the difficulty of all PISA items. Estimates of student ability and item difficulty can then be mapped on a single continuum. The relative ability of students taking a particular test can be estimated by considering the percentage of test items they get correct. The relative difficulty of items in a test can be estimated by considering the percentage of students getting each item correct. In PISA, all students within a level are expected to answer at least half of the items from that level correctly. Students at the bottom of a level are able to provide the correct answers to about 52 percent of all items from that level, have a 62 percent chance of success on the easiest items from that level, and have a 42 percent chance of success on the most difficult items from that level. Students in the middle of a level have a 62 percent chance of correctly answering items of average difficulty for that level (an overall response probability of 62 percent). Students at the top of a level are able to provide the correct answers to about 70 percent of all items from that level, have a 78 percent chance of success on the easiest items from that level, and have a 62 percent chance of success on the most difficult items from that level. Students just below the top of a level would score less than 50 percent on an assessment at the next higher level. Students at a particular level demonstrate not only the knowledge and skills associated with that level but also the proficiencies defined by lower levels. Patterns of responses for students below level 1b for reading literacy and below level 1 for mathematics and science literacy suggest that these students are unable to answer at least half of the items from those levels correctly. For details about the approach to defining and describing the PISA levels and establishing the cut scores, see the OECD's *PISA 2012 Technical Report* (forthcoming). The table on the following page shows the cut scores for each proficiency level for mathematics, science, and reading literacy.

Table A-1. Cut scores for proficiency levels for mathematics, science, and reading literacy: 2012

Proficiency level	Mathematics	Science	Reading[1]
Below level 1	0-357.77	0-334.94	0-262.04
Level 1	greater than 357.77-420.07	greater than 334.94-409.54	greater than 262.04-334.75 (1b) greater than 334.75-407.47 (1a)
Level 2	greater than 420.07-482.38	greater than 409.54-484.14	greater than 407.47-480.18
Level 3	greater than 482.38-544.68	greater than 484.14-558.73	greater than 480.18-552.98
Level 4	greater than 544.68-606.99	greater than 558.73-633.33	greater than 552.98-625.61
Level 5	greater than 606.99-669.30	greater than 633.33-707.93	greater than 625.61-698.32
Level 6	greater than 669.30-1000	greater than 707.93-1000	greater than 698.32-1000

[1]The first reading literacy proficiency level is composed of levels 1a and 1b. The score range for below level 1 refers to scores below level 1b.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Data Limitations

As with any study, there are limitations to PISA 2012 that should be taken into consideration. Estimates produced using data from PISA 2012 are subject to two types of error: nonsampling errors and sampling errors.

Nonsampling error is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. For example, suppose the study was unsuccessful in getting permission from many rural schools in a certain region of the country. In that case, reports of means for rural schools for that region may be biased. Fortunately, such a coverage problem did not occur in PISA in the United States. The sources of nonsampling errors are typically problems such as unit and item nonresponse, the differences in respondents' interpretations of the meaning of survey questions, and mistakes in data preparation.

Sampling errors arise when a sample of the population, rather than the whole population, is used to estimate some statistic. Different samples from the same population would likely produce somewhat different estimates of the statistic in question. This fact means that there is a degree of uncertainty associated with statistics estimated from a sample. This uncertainty is referred to as sampling variance and is usually expressed as the standard error of a statistic estimated from sample data. The approach used for calculating standard errors in PISA was the Fay method of balanced repeated replication (BRR) (Judkins 1990). This method of producing standard errors uses information about the sample design to produce more accurate standard errors than would be produced using simple random sample assumptions.

Standard errors can be used as a measure for the precision expected from a particular sample. Standard errors for all statistics reported in this report are available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024.

Confidence intervals provide a way to make inferences about population statistics in a manner that reflects the sampling error associated with the statistic. Assuming a normal distribution and a 95 percent confidence interval, the population value of this statistic can be inferred to lie within the confidence interval in 95 out of 100 replications of the measurement on different samples drawn from the same population.

Confidentiality and Disclosure Limitations

Confidentiality analyses for the United States were designed to provide reasonable assurance that public-use data files issued by the PISA consortium would not allow identification of individual U.S. schools or students when compared against other public-use data collections. Disclosure limitations included identifying and masking potential disclosure risk to PISA schools and including an additional measure of uncertainty to school and student identification through random swapping of data elements within the student and school file. Swapping was designed to not significantly affect estimates of means and variances for the whole sample or reported subgroups (Krenzke et al. 2006).

Statistical Procedures

Comparisons made in the text of this report have been tested for statistical significance. For example, in the commonly made comparison of OECD averages to U.S. averages, tests of statistical significance were used to establish whether or not the observed differences from the U.S. average were statistically significant.

In almost all instances, the tests for significance used were standard *t* tests. These fell into two categories according to the nature of the comparison being made: comparisons of independent samples and comparisons of nonindependent samples. In PISA, education system groups are independent. We judge that a difference is "significant" if the probability associated with the *t* test is less than .05. If a test is significant this implies that difference in the observed means in the sample represents a real difference in the population.[6] No adjustments were made for multiple comparisons.

In simple comparisons of independent averages, such as the average score of education system 1 with that of education system 2, the following formula was used to compute the *t* statistic:

$$t = \frac{(est_1 - est_2)}{\sqrt{se_1^2 + se_2^2}},$$

where est_1 and est_2 are the estimates being compared (e.g., averages of education system 1 and education system 2) and se_1^2 and se_2^2 are the corresponding squared standard errors of these averages. The PISA 2012 data are hierarchical and include school and student data from the participating schools. The standard errors for each education system take into account the clustered nature of the sampled data. These standard errors are not adjusted for correlations between groups since groups are independent.

The second type of comparison occurs when evaluating differences between nonindependent groups within the education system. Because of the sampling design in which schools and students within schools are randomly sampled, the data within the education system from mutually exclusive sets of students (for example, males and females) are not independent. As a result, to determine whether the performance of females differs from the performance of males, for example, the standard error of the difference taking into account the correlation between females' scores and males' scores needs to be estimated. A BRR procedure, described above, was used to estimate the standard errors of differences

[6] A .05 probability implies that the *t* statistic is among the 5 percent most extreme values one would expect if there were no difference between the means. The decision rule is that when *t* statistics are this extreme, they are sampled from a population where there is a difference between the means.

between nonindependent samples within the United States. Use of the BRR procedure implicitly accounts for the correlation between groups when calculating the standard errors.

To test comparisons between nonindependent groups the following *t* statistic formula was used:

$$t = \frac{(est_{grp1} - est_{grp2})}{se_{(grp1 - grp2)}},$$

where est_{grp1} and est_{grp2} are the nonindependent group estimates being compared and $se_{(grp1-grp2)}$ is the standard error of the difference calculated using BRR to account for the correlation between the estimates for the two nonindependent groups.

This page intentionally left blank

Appendix B: International and U.S. Data Collection Results

This appendix describes the success of participating education systems in meeting the international technical standards on data collection described in appendix A. Information is provided for all participating education systems on their coverage of the target population, exclusion rates, and response rates. This appendix also provides the U.S. response rates and the results of the U.S. nonresponse bias analysis.

Response Rates

Table B-1 provides information on weighted school participation rates before and after school replacement and the number of participating schools after replacement for each participating education system. Table B-2 provides information on coverage of the target population, overall exclusion rates, weighted student response rates after school replacement, and the number of participating students after replacement for each participating education system.

One hundred thirty-nine participating original schools and 23 replacement schools participated in the U.S. administration of the Program for International Student Assessment (PISA), for a total of 162 schools. Although all 162 schools were included in the analysis of the U.S. PISA 2012 results, international guidelines stipulated that schools with between 25 and 50 percent of students participating were considered nonparticipating schools for the purposes of calculating response rates (but were eligible to be included in the analysis of results). In the United States, one replacement school had a student response rate between 25 and 50 percent. This resulted in 161 participating schools and an overall weighted school response rate of 77 percent. The overall weighted student response rate was 89 percent and the U.S. overall student exclusion rate was 5 percent.

In Connecticut, there were 50 participating schools (out of 51 eligible schools), resulting in an overall weighted school response rate of 98 percent. The overall weighted student response rate was 87 percent and the overall student exclusion rate was 4 percent. In Florida, there were 54 participating schools (out of 54 eligible schools), resulting in an overall weighted school response rate of 100 percent. The overall weighted student response rate was 90 percent and the overall student exclusion rate was 8 percent. In Massachusetts, there were 49 participating schools (out of 49 eligible schools), resulting in an overall weighted school response rate of 100 percent. The overall weighted student response rate was 90 percent and the overall student exclusion rate was 4 percent.

Table B-1. Number of schools and weighted participation rates, by education system: 2012

Education system	Percent		
	Weighted school participation before replacement	Weighted school participation after replacement	Number of participating schools after replacement
Albania	100.0	100.0	204
Argentina	95.5	95.9	219
Australia	97.9	97.9	757
Austria	100.0	100.0	191
Belgium	84.4	96.6	282
Brazil	92.7	95.4	837
Bulgaria	99.2	99.8	187
Canada	91.3	92.9	840
Chile	91.9	98.8	221
Chinese Taipei	100.0	100.0	163
Colombia	86.6	97.4	352
Costa Rica	98.9	98.9	191
Croatia	98.7	99.9	163
Cyprus	96.6	96.6	117
Czech Republic	98.1	99.6	295
Denmark	87.0	95.5	339
Estonia	100.0	100.0	206
Finland	99.0	99.3	311
France	96.6	96.6	223
Germany	97.7	98.0	228
Greece	93.2	98.9	188
Hong Kong-China	78.7	94.1	147
Hungary	97.6	99.4	204
Iceland	99.3	99.3	133
Indonesia	94.9	98.0	206
Ireland	98.7	99.3	183
Israel	91.1	93.8	172
Italy	89.1	97.4	1,186
Japan	86.3	95.5	191
Jordan	100.0	100.0	233
Kazakhstan	100.0	100.0	218
Korea, Republic of	99.9	99.9	156
Latvia	87.9	99.9	211
Liechtenstein	100.0	100.0	12
Lithuania	98.2	100.0	216
Luxembourg	100.0	100.0	42
Macao-China	100.0	100.0	45

See notes at end of table.

Table B-1. Number of schools and weighted participation rates, by education system: 2012—Continued

Education system	Percent: Weighted school participation before replacement	Percent: Weighted school participation after replacement	Number of participating schools after replacement
Malaysia	100.0	100.0	164
Mexico	91.8	95.3	1,468
Montenegro, Republic of	100.0	100.0	51
Netherlands	75.3	89.4	177
New Zealand	80.9	89.3	177
Norway	85.2	94.7	197
Peru	97.9	98.6	240
Poland	85.4	97.9	182
Portugal	95.4	95.8	187
Qatar	99.9	99.9	157
Romania	100.0	100.0	178
Russian Federation	100.0	100.0	227
Serbia, Republic of	90.0	95.4	152
Shanghai-China	100.0	100.0	155
Singapore	97.5	98.2	172
Slovak Republic	87.5	99.0	231
Slovenia	98.1	98.1	335
Spain	99.7	99.7	902
Sweden	98.9	99.8	209
Switzerland	94.5	98.3	410
Thailand	98.0	100.0	239
Tunisia	99.3	99.3	152
Turkey	97.5	99.9	169
United Arab Emirates	99.4	99.4	453
United Kingdom	80.1	89.2	505
United States	**67.1**	**77.2**	**161**
Uruguay	99.4	100.0	180
Vietnam	100.0	100.0	162
U.S. state education systems			
Connecticut	98.0	98.0	50
Florida	100.0	100.0	54
Massachusetts	100.0	100.0	49

NOTE: In calculating school participation rates, each school received a weight equal to the product of its base weight (the reciprocal of its probability of selection) and the number of age-eligible students enrolled in the school, as indicated on the sampling frame. Weighted school participation before replacement refers to the sum of weights of the original sample schools with PISA-assessed students and a student response rate of at least 50 percent over the sum of weights of all original sample schools. Weighted school participation after replacement refers to the sum of weights of the original and replacement schools with PISA-assessed students and a student response rate of at least 50 percent over the sum of weights of responding original sample schools, responding replacement schools, and eligible refusing original sample schools. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public schools only.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

Table B-2. Coverage of target population, student exclusion and weighted participation rates, and number of students, by education system: 2012

Education system	Percent: Coverage of national desired population	Percent: Overall student exclusion rate	Percent: Weighted student participation after replacement	Number of participating students
Albania	99.9	0.1	92.5	4,743
Argentina	99.3	0.7	88.0	5,908
Australia	96.0	4.0	86.8	17,774
Austria	98.7	1.3	91.7	4,756
Belgium	98.6	1.4	90.9	9,690
Brazil	98.6	1.4	90.1	20,091
Bulgaria	97.4	2.5	95.7	5,282
Canada	93.6	6.4	80.8	21,548
Chile	98.7	1.3	94.6	6,857
Chinese Taipei	98.8	1.2	96.3	6,046
Colombia	99.9	0.1	93.1	11,173
Costa Rica	100.0	0.0	89.0	4,602
Croatia	97.8	2.2	92.2	6,153
Cyprus	96.7	3.3	93.3	5,078
Czech Republic	98.2	1.8	90.1	6,535
Denmark	93.8	6.1	89.1	7,481
Estonia	94.2	5.7	92.9	5,867
Finland	98.1	1.9	90.7	8,829
France	95.6	4.3	89.5	5,682
Germany	98.5	1.5	93.2	5,001
Greece	96.4	3.6	96.7	5,125
Hong Kong-China	98.2	1.8	93.1	4,670
Hungary	97.4	2.6	92.7	4,810
Iceland	96.2	3.8	84.7	3,508
Indonesia	99.7	0.3	95.2	5,622
Ireland	95.5	4.5	84.1	5,016
Israel	95.9	4.1	90.0	6,061
Italy	96.7	3.3	92.8	38,142
Japan	97.9	2.1	96.1	6,351
Jordan	99.6	0.4	95.0	7,038
Kazakhstan	96.6	3.3	98.9	5,808
Korea, Republic of	99.2	0.8	98.7	5,033
Latvia	96.0	3.9	90.9	5,276
Liechtenstein	95.8	4.2	93.3	293
Lithuania	96.0	4.0	92.1	4,618
Luxembourg	87.2	8.3	95.2	5,260
Macao-China	99.8	0.2	99.4	5,335
Malaysia	99.8	0.2	94.0	5,197
Mexico	99.3	0.7	93.9	33,806

See notes at end of table.

Table B-2. Coverage of target population, student exclusion and weighted participation rates, and number of students, by education system: 2012—Continued

Education system	Percent: Coverage of national desired population	Percent: Overall student exclusion rate	Percent: Weighted student participation after replacement	Number of participating students
Montenegro, Republic of	99.7	0.3	93.8	4,744
Netherlands	95.6	4.3	85.0	4,460
New Zealand	95.4	4.6	84.7	5,248
Norway	93.9	6.1	90.9	4,686
Peru	99.8	0.2	96.0	6,035
Poland	95.4	4.6	87.6	5,662
Portugal	98.4	1.6	87.4	5,722
Qatar	97.5	2.5	99.7	10,966
Romania	96.5	3.4	97.8	5,074
Russian Federation	97.6	2.4	97.3	6,418
Serbia, Republic of	97.1	2.8	93.4	4,684
Shanghai-China	98.5	1.5	98.5	6,374
Singapore	98.8	1.2	94.3	5,546
Slovak Republic	97.1	2.9	93.8	5,737
Slovenia	98.4	1.6	90.5	7,229
Spain	95.7	4.3	89.9	25,335
Sweden	94.6	5.4	92.2	4,739
Switzerland	95.8	4.1	92.0	11,234
Thailand	98.7	1.3	98.9	6,606
Tunisia	99.8	0.2	90.3	4,407
Turkey	98.5	1.5	98.2	4,848
United Arab Emirates	97.9	2.1	94.7	11,500
United Kingdom	94.6	5.4	86.1	12,659
United States	**94.6**	**5.3**	**88.9**	**6,111**
Uruguay	99.7	0.3	90.0	5,315
Vietnam	99.3	0.7	99.9	4,959
U.S. state education systems				
Connecticut	95.9	4.1	87.5	1,697
Florida	91.7	8.3	90.0	1,896
Massachusetts	95.6	4.4	90.0	1,723

NOTE: In calculating student participation rates, each student received a weight (student base weight) equal to the product of the school base weight—for the school in which the student was enrolled—and the reciprocal of the student selection probability within the school. Coverage of 15-year-old population refers to the extent to which the weighted participants covered the target population of all enrolled students in grades 7 and above. Coverage of national desired population refers to the extent to which the weighted participants covered the national population of 15-year-olds under the nonexcluded portion of the student sample. Overall student exclusion rate is the percentage of students excluded for intellectual or functional disabilities, or insufficient assessment language experience at either the school level or within schools. Weighted student participation after replacement refers to the sum of weights of students in original and replacement schools with PISA-assessed students and a student response rate of at least 50 percent over the sum of weights of students in responding original sample schools, responding replacement schools, and eligible refusing original sample schools. Italics indicate non-OECD countries and education systems. Results for Connecticut, Florida, and Massachusetts are for public school students only.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2012.

U.S. Nonresponse Bias Analysis

Of the 240 original sampled schools in the U.S. national sample, 207 were eligible (18 schools did not have any 15-year-olds enrolled, 6 had closed, and 9 were otherwise ineligible), and 139 agreed to participate. The weighted school response rate before replacement was 67 percent, requiring the United States to conduct a nonresponse bias analysis, which was used by the PISA consortium and the Organization for Economic Cooperation and Development (OECD) to evaluate the quality of the final sample.

A bias analysis was conducted in the United States to address potential problems in the data owing to school nonresponse. To compare PISA participating schools to the total eligible sample of schools, it was necessary to match the sample of schools to the sample frame to identify as many characteristics as possible that might provide information about the presence of nonresponse bias. Frame characteristics were taken from the 2008–09 Common Core of Data for public schools and from the 2009–10 Private School Universe Survey for private schools. The available school characteristics included affiliation (public or private), locale (city, suburb, town, rural), Census region, number of age-eligible students, total number of students, and percentage of various racial/ethnic groups (White, Black, Hispanic, non-Hispanic, Asian, American Indian or Alaska Native, Native Hawaiian/Pacific Islander, and multiracial). The percentage of students eligible for free or reduced-price lunch was available for public schools only. The full text of the nonresponse bias analysis conducted for PISA 2012 will be included in the technical report released with the U.S. national dataset (Kastberg, Roey, Lemanski, Chan, and Murray forthcoming).

For original sample schools, participating schools had a higher mean percentage of Hispanic students than the total eligible sample of schools (21.1 versus 18.1 percent, respectively). Participating original sample schools also had a higher mean percentage of students eligible for free or reduced-price lunch than did the total eligible sample of schools (39.3 versus 36.1 percent, respectively). When all factors were considered simultaneously in a logistic regression analysis, only "town" (a territory inside an urban cluster with a core population between 25,000 and 50,000) was a significant predictor of participation. The percentage of students eligible for free or reduced-price lunch was not included in the logistic regression analysis as public and private schools were modeled together using only the variables available for all schools.[1]

For final sample schools (with substitutes), participating schools had a higher mean percentage of students eligible for free or reduced-price lunch than the total eligible sample of schools (38.4 versus 36.2 percent, respectively). When all factors were considered simultaneously in a logistic regression analysis (again with free or reduced-price lunch eligibility omitted), no variables were statistically significant predictors of participation.

With the inclusion of substitute schools and school nonresponse adjustments applied to the weights, only the percentage of students eligible for free or reduced-price lunch remained statistically significant. Specifically, the participating schools had a higher mean percentage of students eligible to receive free or reduced-price lunch than the total eligible sample of schools (38.4 versus 36.2 percent, respectively). However, there was not a statistically significant relationship between participating schools and the total frame of eligible schools for the percentage of students eligible for free or reduced-price lunch (38.4 versus 37.1 percent, respectively). We therefore conclude that, despite

[1] The nonresponse bias analysis was designed to measure the potential nonresponse bias for all participating schools, so no additional logistic regression was conducted using only public schools.

the tendency of schools with higher percentages of students eligible for free and reduced-price lunch to participate at a greater rate than other sampled schools, there is little evidence of resulting potential bias in the final sample. The multivariate regression analysis cannot be conducted after the school nonresponse adjustments are applied to the weights. The concept of nonresponse adjusted weights does not apply to the nonresponding units, and, thus, we cannot conduct an analysis that compares respondents with nonrespondents using nonresponse adjusted weights.

In sum, the investigation into nonresponse bias at the school level in the United States in PISA 2012 provides evidence that there is little potential for nonresponse bias in the PISA participating sample based on the characteristics studied. It also suggests that, while there is little evidence that the use of substitute schools reduced the potential for bias, it has not added to it. Moreover, the application of school nonresponse adjustments substantially reduced the potential for bias.

www.ed.gov

ies.ed.gov